

COMPOUND HETEROZYGOUS MISSENSE AND DEEP INTRONIC VARIANTS IN *NDUFAF6* UNRAVELED  
BY EXOME SEQUENCING AND mRNA ANALYSIS

Alessia Catania<sup>1</sup>, Anna Ardisson<sup>2</sup>, Daniela Verrigni<sup>3</sup>, Andrea Legati<sup>1</sup>, Aurelio Reyes<sup>4</sup>, Eleonora Lamantea<sup>1</sup>, Daria Diodato<sup>3</sup>, Davide Tonduti<sup>2</sup>, Valentina Imperatore<sup>6</sup>, Anna Maria Pinto<sup>6,7</sup>, Isabella Moroni<sup>2</sup>, Enrico Bertini<sup>3</sup>, Alan Robinson<sup>4</sup>, Rosalba Carrozzo<sup>3</sup>, Massimo Zeviani<sup>4</sup>, and Daniele Ghezzi<sup>1,8</sup>

*1- Molecular Neurogenetics, Foundation IRCCS Neurological Institute Besta, Milan, Italy*

*2- Child Neurology Unit, Foundation IRCCS Neurological Institute Besta, Milan, Italy*

*3- Department of Molecular and Translational Medicine DIMET, University of Milan-Bicocca, Milan, Italy*

*4- Unit of Muscular and Neurodegenerative Disorders, Laboratory of Molecular Medicine, "Bambino Gesù" Children's Hospital, IRCCS, Rome, Italy*

*5- Medical Research Council - Mitochondrial Biology Unit, University of Cambridge, Cambridge, UK*

*6- Medical Genetics, Department of Medical Biotechnology, University of Siena, Policlinico "Santa Maria alle Scotte", Siena, Italy*

*7- Medical Genetics, Azienda Ospedaliero-Universitaria Senese, Siena, Italy*

*8- Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy*

Correspondence to: Daniele Ghezzi

Molecular Neurogenetics Unit, Foundation IRCCS Neurological Institute "Besta", via Temolo 4,  
20126 Milan, Italy; Tel: +39 0223942627; Fax: +39 0223942619; daniele.ghezzi@istituto-besta.it

**Running title:**

MISSENSE AND INTRONIC VARIANTS IN *NDUFAF6*

## Abstract

Biallelic mutations in *NDUFAF6* have been identified as responsible for cases of autosomal recessive Leigh syndrome associated with mitochondrial complex I deficiency. Here, we report two siblings and two unrelated subjects with Leigh syndrome, in which we found the same compound heterozygous missense (c.532G>C:p.A178P) and deep intronic (c.420+784C>T) variants in *NDUFAF6*. We demonstrated that the identified intronic variant creates an alternative splice site, leading to production of an aberrant transcript.

A detailed analysis of whole exome sequencing data together with functional validation based on mRNA analysis may reveal pathogenic variants even in non-exonic regions.

## 1. Introduction

Leigh syndrome is an early onset progressive neurodegenerative disease with an invariably devastating clinical course. The neuropathology of Leigh syndrome is characterized by bilateral, symmetrical necrotic lesions in deep gray matter structures. The most common cause of Leigh syndrome is a defect in oxidative phosphorylation.<sup>1</sup> The NADH dehydrogenase (ubiquinone) complex I assembly factor 6 (*NDUFAF6*, previously known as *C8ORF38*) gene encodes a mitochondrial protein, highly conserved across eukaryotes, which plays an essential role in the early assembly stages of mitochondrial respiratory chain complex I.<sup>2</sup> Biallelic missense mutations within *NDUFAF6* have been associated with cases of Leigh syndrome due to mitochondrial complex I deficiency.<sup>3-5</sup> In a very recent paper about molecular and enzymatic assays for diagnosis in Leigh syndrome cases, *NDUFAF6* was one of the most frequently mutated nuclear genes and was invariably associated with reduced complex I activity.<sup>6</sup>

Recently, Bianciardi et al.<sup>7</sup> described a patient presenting Leigh syndrome and mitochondrial complex I deficiency associated with a pathogenic heterozygous missense variant in *NDUFAF6* (c.532G>C;p.A178P) and an almost monoallelic expression of the mutated allele at transcriptional level; an autosomal recessive model of inheritance was hypothesized, but the second pathogenic mutation remained unidentified.<sup>7</sup> Upon identification in our laboratory of two siblings and a singleton unrelated subject, all affected by Leigh syndrome and harbouring the same heterozygous missense mutation (c.532G>C;p.A178P), here we provide evidence that the second allelic mutation consists of a deep intronic variant present in all affected individuals, including the previously published case.<sup>7</sup> Through mRNA analysis we demonstrated that the identified intronic mutation is responsible for the formation of an alternative splice site, leading to altered mRNA splicing.

## 2. Materials and Methods

Informed consents for biological sample collection and genetic studies were obtained from all the subjects involved in the study (Siblings A1 and A2 and their parents; patient B, previously described by Bianciardi et al. <sup>7</sup>; patient C and his parents), in agreement with the Declaration of Helsinki.

Extraction of genomic DNA from peripheral blood was performed using standard methods. A targeted-NGS using a custom panel containing genes responsible for mitochondrial disorders<sup>8</sup> was performed on DNAs from patients A1 and B, while clinical exome sequencing using a commercial kit with genes associated with inherited diseases (TrusightOne, Illumina) was performed in patient C. Whole exome sequencing (WES) analysis and variant calling/prioritization were conducted on DNAs from patients A1 and A2, as previously described.<sup>8</sup> Libraries were prepared with the Nextera Rapid Exome Capture kit (Illumina), and sequencing was performed on an Illumina MiSeq platform. Variants identified by NGS were validated by Sanger sequencing and resolved on a 3130xl Genetic Analyzer (Applied Biosystems).

The identified mutations were validated by employing the following pairs of primers, respectively:

Exon 3b F: 5'-ATTTTGAGGTGGTGATTTCAAAG-3'; R: 5'-GGAAGGCTTTAGTGGTAAACTGG-3';

Exon 5 F: 5'-AATGGAAACCTATGGGCTAGAG-3'; R: 5'-TGAGCTTCTTGAAGTGGGATG-3'.

For RNA purification and cDNA retrotranscription we used RNeasy mini kit (QIAGEN) and GoTaq 2-Step RT-qPCR System (Promega), respectively, according to the manufacturers' protocols.

For cDNA amplification (from exon 3 to exon 7) and sequencing we used the following primer pair:

3F: 5'-CTGAGAAAACAATTGGACTGATG-3'; 7R: 5'-TCTCACATTTTATCTTGGTTCCTC-3'.

For deep sequencing of *NDUFAF6* transcript, PCR products were processed with Nextera XT DNA sample preparation kit (Illumina). Next Generation Sequencing (NGS) was performed on an Illumina MiSeq instrument. Filtered reads were then aligned using BWA and visualized with Integrating Genomics Viewer (IGV).

Tag SNPs were selected using the SNPinfo web server (<https://snpinfo.niehs.nih.gov>), based on the HapMap CEU (Utah residents with Northern and Western European ancestry) and Tuscani populations. Specific primers were designed (available upon request) and SNPs were assessed by Sanger sequencing. Information about SNPs frequencies was obtained from public databases: dbSNP, Genome Aggregation Database (gnomAD).

For the nonsense-mediated mRNA decay (NMD) assay, patients and control fibroblasts were grown in complete DMEM and incubated for 12 h with puromycin (100 µg/ml). Total RNA was extracted after 4 and 8 hours of recovery in puromycin free medium from treated and untreated paired cultures.

### **3. Results**

#### *3.1. Case reports*

We report an Italian family in which two siblings were diagnosed with Leigh syndrome (Family A, Fig. 1A). The older sibling (patient A1) presented psychomotor regression at 21 months of age; neurological evaluation at 30 months showed ataxic gait and fine tremor. Cognitive functions were preserved. He presented a Leigh syndrome pattern at brain MRI. Biochemical analysis showed an isolated respiratory chain complex I deficiency in muscle and fibroblasts (42% and 38% of the controls' mean, respectively; Suppl. Table S1). The disease was slowly progressive: at 4½ years he presented with drooling, dysarthria, dysmetria, tremor, severe ataxic gait and hypertonia. His younger sister (patient A2) showed psychomotor delay at the age of 12 months and then limb dysmetria, trunk titubation, and ataxic gait. Cognitive functions were preserved. MRI disclosed a Leigh syndrome pattern. Plasma lactate and pyruvate levels were slightly increased. A detailed clinical description of these cases is reported in the supplementary material.

Patient B was described in detail by Bianciardi et al.<sup>7</sup> Briefly, he developed motor and language disturbances by the age of 3.5 years progressing to severe gait impairment, dysarthria and early occurrence of dystonic movements. Brain MRI showed necrotic damage of the putamina, gradually extending to dentate nuclei and anterior caudate nuclei. Evaluation of mitochondrial respiratory chain revealed decreased complex I activity on fibroblasts (~25% of the controls' mean), while enzymatic activity was normal on muscle homogenate (Suppl. Table S1).

Patient C is an 11 year-old Italian boy, first child of unrelated parents. Psychomotor developmental milestones were referred as normal. He presented at 5 years of age with gait unsteadiness and motor coordination problems. These symptoms gradually worsened, configuring over time an extrapyramidal syndrome; cognitive function were preserved. Brain MRI disclosed involvement of the putamen bilaterally, stable over time. Extensive metabolic screening was normal as well as mtDNA analysis. A detailed clinical description of this case is reported in the supplementary material.

### *3.2. Molecular studies*

WES was performed on patients A1 and A2. Variants annotation and filtering steps focused on variants shared by both siblings and affecting mitochondrial genes led to the identification of the heterozygous missense variant c.532G>C:p.A178P in *NDUFAF6* (NM\_152416.3) in both siblings; the same mutation (rs201088736) had been already reported as pathogenic by Bianciardi et al.<sup>7</sup> Analysis of WES data also revealed the presence of an additional heterozygous intronic variant (NM\_152416.3:c.420+784C>T) in *NDUFAF6* in both affected siblings (Fig. 1A-B). It corresponds to rs749738738 in dbSNP database, and is reported with an extremely low allele frequency (0.003% in gnomAD). The missense and the intronic variants, validated by Sanger sequencing, were inherited from the unaffected father and mother, respectively (Fig. 1A).

In order to assess the effects of the intronic variant, RNA was extracted from fresh blood samples of family A members and retrotranscribed into cDNA. Through amplification and sequencing of the *NDUFAF6* transcript (Fig. 1C-D), we detected an additional transcript in samples of both siblings and their mother (Suppl. Fig. S1): this corresponded to alternative splicing isoforms (e.g. ENST00000520757.1 or XM\_005250791.1), retaining an extra 124 nucleotide-long exon (exon 3b; Fig. 1B) and predicted to undergo premature non-sense mediated decay on public databases, because of the creation of a premature stop codon. Interestingly, the alternative splicing acceptor site is placed 11 nucleotides downstream the rs749738738 variant and *in silico* predictions gave higher scores for creating an acceptor site to the mutant sequence (Suppl. Table S2).

We also assessed and eventually confirmed the presence of the same intronic variant on DNA of patient B, already reported to be heterozygous for the paternally inherited c.532G>C:p.A178P mutation.<sup>7</sup> In accordance with a recessive pattern of transmission, segregation analysis confirmed a maternal origin for the intronic variant.

Similarly, direct sequencing revealed the c.420+784C>T intronic variant in an additional case, patient C, in which an independent genetic analysis using a clinical exome panel had identified the c.532G>C:p.A178P heterozygous variant. The two variants were inherited from the parents (Fig. 1). Increased expression of the alternative splicing isoform of the transcript (retaining exon 3b) was confirmed on RNA purified from patient B's fibroblasts as well.

A further analysis of *NDUFAF6* transcripts using an NGS approach confirmed that, although expressed at low levels also in control subjects, the alternative isoform is overrepresented in the samples from individuals harboring the intronic variant (patients A1 and B, and the mother of siblings A1 and A2) compared to subjects lacking this variant (control subjects and the father of siblings A1 and A2) (Fig. 2).

Notably, in the patients we observed a predominant expression of the mutated allele harboring the missense mutation c.532G>C in fibroblasts and, partly, in blood (Fig. 1D, Fig. 2). A non-sense



mediated decay (NMD) assay performed on fibroblasts from both patients A1 and B and a control individual did not show any detectable difference in RNA sequence profiles between treated and untreated samples.

The identification of two extremely rare variants in three unrelated families from Italy displaying a similar clinical presentation raised the possibility of unreported/unknown relationship between the families and common founder alleles. All the families originate from Southern Italy, although from different regions (2 from Campania, 1 from Abruzzo). We performed a haplotype analysis using 7 tagSNPs present in the *NDUFAF6* genomic region. All the patients showed the same haplotype (Suppl. Table S3) indicating that they likely harbor common *NDUFAF6* alleles rather than having had independent mutational events. Analysis of the variants identified by targeted-NGS confirmed this finding, with shared *NDUFAF6* haplotypes amongst patients; however it revealed a quite different variants profile in surrounding genes present on chromosome 8 (Suppl. Table S4), suggesting not recent founder occurrence.

#### **4. Discussion**

We identified compound heterozygous missense and deep intronic variants in three different families with a typical recessive *NDUFAF6*-related disease characterized by Leigh syndrome. Complementation studies, previously performed on patient B,<sup>7</sup> already proved that *NDUFAF6* impairment was the cause of the disease. In accordance with *in silico* predictions, our transcript data suggest that the intronic variant promotes the alternative splicing event, thus enhancing expression of the non-functional altered isoform of the gene at the expense of the canonical one.

An NMD assay, together with the amplification of the aberrant transcript in standard conditions, excludes the hypothesis of a relevant NMD involvement on the alternative *NDUFAF6* isoform, at least in fibroblasts. Nevertheless, several *NDUFAF6* transcripts which could be variably expressed in

different tissues have been reported<sup>2</sup> and may be differently affected by the intronic variant and otherwise subjected to NMD.

In line with our data, Bianciardi et al, have previously shown an almost exclusive expression of the missense exonic mutation in blood and fibroblasts-derived RNA samples from Patient B in comparison with urine and brush RNA samples which showed a heterozygous expression for this variant. These data could thus suggest the existence of trans-splicing factors which determine a tissue-specific switch between the alternative spliced isoform and the canonical isoform, with the alternative spliced isoform being highly expressed in tissues such as blood and fibroblasts rather than urotelium (Figure 3). According to this hypothesis, in blood and fibroblasts from our patients, due to the presence of the intronic variant, the predominant inclusion of an alternative exon in about a 50% of the total transcript would lead to a highly unstable not detectable isoform and consequently to the apparent homozygous expression of the missense exonic mutation. The fact that we identified three unrelated cases with the same, peculiar combination of extremely rare *NDUFAF6* variants may be casual. Otherwise we may speculate that one of these variants is hypomorphic (possibly the splicing variant which allows the formation of a quote of wild type transcript) and does not lead to disease, at least Leigh syndrome, in the homozygous state but only when associated with a severe mutation which in turn is too deleterious/embryolethal in the homozygous state. Homozygosity or various compound heterozygosity of splicing defect, missense changes or nonsense mutations have been found in the previously reported patients with *NDUFAF6* mutations<sup>3-6</sup>, and the affected nucleotides are located throughout the gene hampering an easy identification of obvious genotype-phenotype correlations.

The clinical presentation of the *NDUFAF6*-mutant cases here reported is a Leigh syndrome with early childhood onset (1-5 years) and stable or slowly progressive course, associated with isolated cerebellar or extrapyramidal signs depending on affected cerebral structures; cognitive involvement

and seizures were not detected/reported. MRI pattern was characterized by striatal and dentate nuclei involvement. Patient C presented isolated putamina involvement that was stable over time, and a relatively mild phenotype compared to other patients. Only in patient A2 putamen alterations were not observed, but she performed a single MRI at disease onset and additional data about radiological course were not available.

Reports on mutations located outside the coding regions and associated with human diseases are rapidly growing. In this study we confirmed that WES data analysis should take into account all rare variants, including those with poorly predictable effect on transcript/protein. Moreover we underline that mRNA analysis is a complementary strategy, extremely useful to integrate WES and demonstrate the deleterious impact of identified variants, especially those affecting splicing and stability of transcripts.

## **Acknowledgments**

This work was supported by the Telethon Grant GGP15041; the Pierfranco and Luisa Mariani Foundation (to DG); the MRC-QQR (2015-2020) grant; the ERC advanced grant FP7-322424, the NRJ-Institut de France grant (to MZ); the Italian Ministry of Health (Ricerca Corrente 2016-2017 to DG, EB and RC).

The “Cell lines and DNA Bank of Genetic Movement Disorders and Mitochondrial Diseases” of the Telethon Network of Genetic Biobanks (grant GTB12001J) and the EuroBioBank Network supplied biological specimens.

**Supplementary information** is available at Journal of Human Genetics website.

## **Conflicts of interest**

The authors declare no conflicts of interests.

## References

- [1] Dahl, H.-H. Getting to the nucleus of mitochondrial disorders: identification of respiratory chain-enzyme genes causing Leigh syndrome. *Am. J. Hum. Genet.* **63**, 1594-1597 (1998).
- [2] M. McKenzie, E.J. Tucker, A.G. Compton, M. Lazarou, C. George, D.R. Thorburn, M.T. Ryan, Mutations in the gene encoding C8orf38 block complex I assembly by inhibiting production of the mitochondria-encoded subunit ND1. *J. Mol. Biol.* **414**, 413-426 (2011).
- [3] Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S.-E., et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-123 (2008).
- [4] Kohda, M., Tokuzawa, Y., Kishita, Y., Nyuzuki, H., Moriyama, Y., Mizuno, Y., et al. A comprehensive genomic analysis reveals the genetic landscape of mitochondrial respiratory chain complex deficiencies. *PLoS Genet.* **12**, e1005679 (2016).
- [5] Fang, F., Liu, Z., Fang, H., Wu, J., Shen, D., Sun, S., et al. The clinical and genetic characteristics in children with mitochondrial disease in China. *Sci. China Life Sci.* **60**, 746-757 (2017).
- [6] Ogawa, E., Shimura, M., Fushimi, T., Tajika, M., Ichimoto, K., Matsunaga, A., et al. Clinical validity of biochemical and molecular analysis in diagnosing Leigh syndrome: a study of 106 Japanese patients. *J. Inherit. Metab. Dis.* **40**, 685-693 (2017).
- [7] Bianciardi, L., Imperatore, V., Fernandez-Vizarra, E., Lopomo, A., Falabella, M., Furini, S., et al. Exome sequencing coupled with mRNA analysis identifies NDUFAF6 as a Leigh gene. *Molec. Genet. Metab.* **119**, 214-222 (2016).
- [8] Legati, A., Reyes, A., Nasca, A., Invernizzi, F., Lamantea, E., Tiranti, V., et al. New genes and pathomechanisms in mitochondrial disorders unraveled by NGS technologies. *Biochim. Biophys. Acta* **1857**, 1326-1335 (2016).

## Figure Legends

### Figure 1. Pedigrees and genetic studies

A. Pedigrees of the families A, B, C showing the identified *NDUFAF6* variants. Black symbols indicate the affected siblings.

B. Schematic structure of *NDUFAF6* NM\_152416.3 transcript (upper panel) and alternative splicing isoform (bottom panel). The retention of a string of 124 nucleotides located within intron 3 (exon 3b, orange box) generates a premature stop codon within exon 4 (arrow).

C. Sanger sequence of *NDUFAF6* cDNA obtained from patient A1 blood RNA, showing two distinct sequences starting from the end of exon 3, that correspond to the NM\_152416.3 transcript (junction exon 3-exon 4) and an alternative transcript retaining exon 3b (junction exon 3-exon 3b).

D. Sanger sequence of *NDUFAF6* cDNA obtained from blood samples of patient A1 (bottom panel) and his father (I-1, upper panel), showing a region within exon 5 which encompasses the heterozygous missense mutation.

### Figure 2. NGS coverage analysis

A. Profiles of depth of coverage obtained through NGS on *NDUFAF6* transcript PCR products. RNA was extracted from lymphocytes of Patient A1 (Pt. A1) and his parents (I-1; I-2), and from fibroblasts of patients A1, B and a control sample (CTR).

B. Quantification of exon 3b retention. Graphs of the coverage ratio (in percentage) between the *NDUFAF6* alternative isoform containing exon 3b and the canonical isoform, calculated for all the samples analyzed by NGS (the two bars correspond to the junctions exon 3-exon 3b and exon 3b-exon 4).

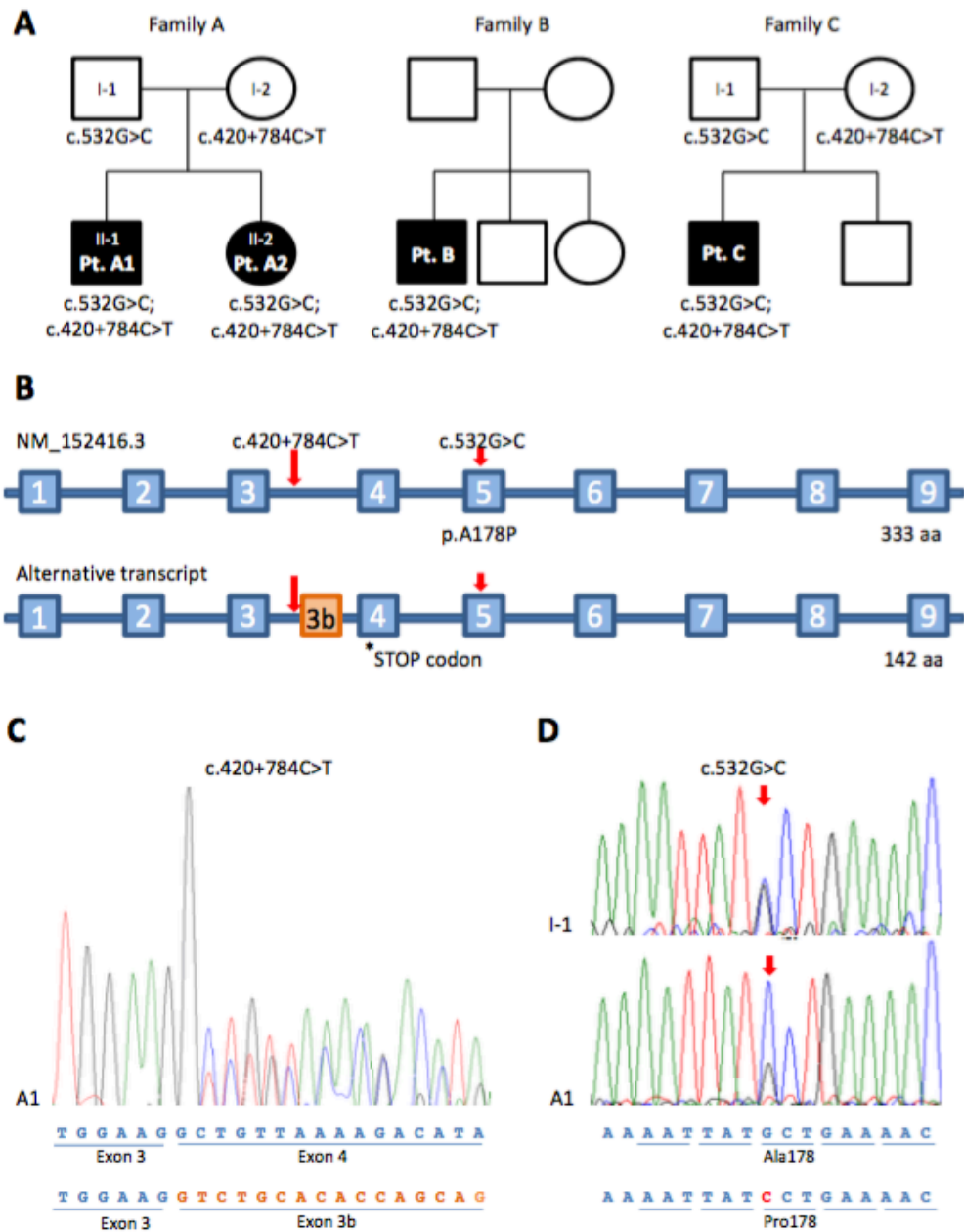
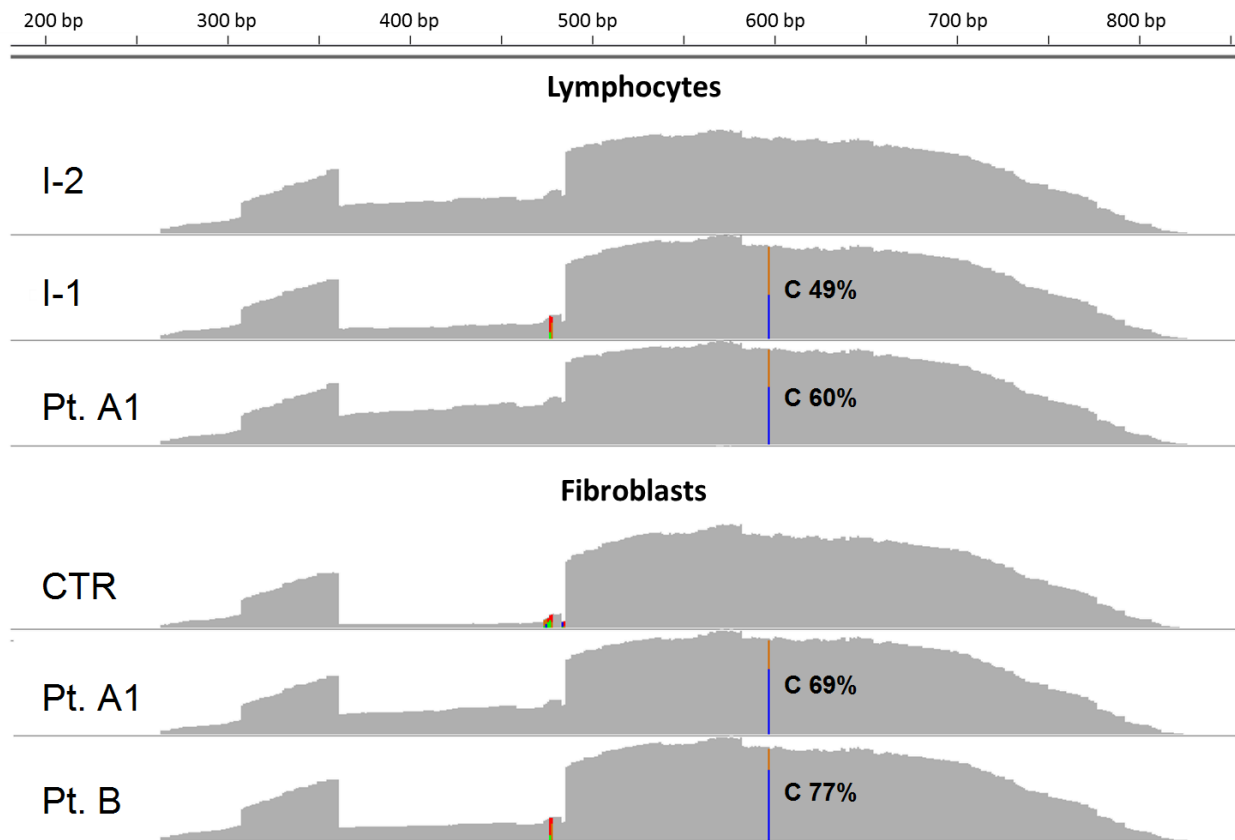
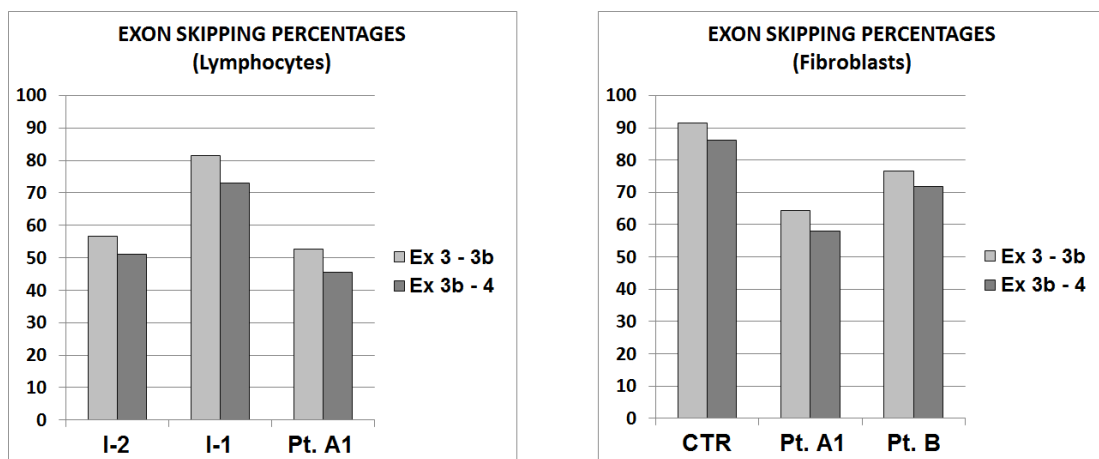


Figure 1

**A****B****Figure 2**